Assessing the Potential of Ride-Sharing Using Mobile and Social Data

Blerim Cici, Athina Markopoulou

Nikolaos Laoutaritis, Enrique Frias-Martinez
Car Usage and Impact

- In USA*:
  - Commuters: 132.3M
  - Driving alone: 79.9%

- Impact
  - Pollution
  - Traffic
  - High car expenses

Introducing Ride-Sharing
An Old Idea, yet …

- Challenges:
  - Live/Work close by
  - Similar schedules
  - Trust among participants

- Opportunities:
  - Smartphones
  - Social media
An Old Idea, yet ...

- Challenges:
  - Live/Work close by
  - Similar schedules
  - Trust among participants

- Opportunities:
  - Smartphones
  - Social media

Full potential of ride-sharing is still unknown
Related Work

Goal: Assess Ride-Sharing Potential

- Q: How many cars can be removed?

- Ideal Data:
  - For all people in a city
  - Full commuting trajectories
  - Willingness to share a ride

- Available Mobile and Social Datasets:
  - Large (but not entire) population
  - Samples of trajectories
  - (Parts of) social media graphs
Goal: Assess Ride-Sharing Potential

- Q: How many cars can be removed?

Ideal Data:
- For all people in a city
- Full commuting trajectories
- Willingness to share a ride

Available Mobile and Social Datasets:
- Large (but not entire) population
- Samples of trajectories
- (Parts of) social media graphs

Find an upper bound to the ride-sharing potential
Outline

- Introduction
- Datasets
- Algorithms for Matching Users
- Results
Call Description Records (CDRs)

- **Spatio-temporal:**
  - Cell tower coordinates
  - Timestamps

- **Social:**
  - Calls among users

- **Details:**
  - Sept – Dec 2009
  - Madrid: 820M calls, 5M users
  - Barcelona: 465M calls, 2M users
Geo-tagged Tweets

- **Spatio-temporal:**
  - (lat,lng) coordinates
  - Timestamps

- **Social:**
  - Twitter Graph

- **Details:**
  - Nov ‘12 – Feb ‘13
  - New York: 5.20M geo-tweets, 225K users
  - Los Angeles: 3.23M geo-tweets, 155K users
Learning from Data 1: Home/Work Locations

- **Methodology**
  - Based on:
  - Ground truth (known home/work):
    - CDRs: Known industrial and residential areas
    - Geo-tweets: Foursquare
  - Train classifiers to identify home/work

- **Home and Work locations inferred:**
  - Madrid (CDRs): 272,479 (out of 5M)
  - NY (Twitter): 71,977 (out of 225K)

- **Home and Work distribution is NOT uniform**
  - In contrast to related work:
Assessing the Potential of Ride-Sharing Using Mobile and Social Data

- Exploit consecutive Home-Work calls
- Home-Work travel
  - Time: Online maps
- Similar for work departure times
\[ d(v, u) = \begin{cases} 
  h(v, u) + w(v, u), & \text{IF } \max(h(v, u), w(v, u)) \leq \delta \\
  \max(|LH(u) - LH(v)|, |LW(u) - LW(v)|) \leq \tau \\
  \infty, & \text{otherwise}
\end{cases} \]

Distance Function
Problem Formulation

- Capacitated Facility Location with Unsplittable Demands:
  - Users: $V$
    - Drivers (facilities): $S \subseteq V$
    - Passengers (clients): $V - S$
  - Capacity: 4 users/car
  - $p(v)$: penalty function for drivers
  - Find:
    - Assignment $a$: $(V - S) \rightarrow S$
    - Minimize:
      $$\sum_{u \in V} d(a(u), u) + \sum_{v \in S} p(v)$$
      - driver-passenger distances
      - driver penalties
Algorithm: EndPoints RS

- Heuristic solution:
  - Based on:
  - Initial solution:
    - b-matching (instead of random)
  - Iterative improvements
    - Scalability
      - Fixed local search steps
      - Fixed numbers of iterations
  - Complexity
    - $O(n \log n)$ for initial solution
    - $O(n)$ to evaluate solution
EndPoint RS for Madrid-CDRs

Success of end-point ride-sharing

- Loose upper bound
- Tighter upper bound
- Time indifferent
- $\tau = 10, \sigma = 20$
- $\tau = 10, \sigma = 30$
- Uniform home/work

Space and time constraints only

Stricter time constraints

Uniform distribution

#users/4

(#users with >1 options)/4

Assessing the Potential of Ride-Sharing Using Mobile and Social Data
Assessing the Potential of Ride-Sharing Using Mobile and Social Data

EndPoint RS for Madrid-CDRs

Success of end-point ride-sharing

- SD of departure time: 30 min
- Distance tolerance: 1 km
- Delay tolerance: 10 min

24% of the cars can be removed!

- #users/4
- (#users with >1 options)/4

Uniform distribution

Assessing the Potential of Ride-Sharing Using Mobile and Social Data
Algorithm: EnRoute RS

- Home/Work paths:
  - Popular Online Maps

- EnRoute RS:
  - Get the solution of EndPoints RS
  - Iterative improvements
  - Fill empty seats by pick-ups

- Spatio-temporal constr. intermediate points:
  - Same and point constraints
Algorithm: EnRoute RS

- Home/Work paths:
  - Popular Online Maps
- EnRoute RS:
  - Get the solution of EndPoints RS
  - Iterative improvements
  - Fill empty seats by pick-ups

Spatio-temporal constr. intermediate points:
  - Same and point constraints

- SD of departure time: 30 min
- Distance tolerance: 1 km
- Delay tolerance: 10 min

53% of the cars can be removed!
Learning from Data 3: Social Ties

- **CDRs graph:**
  - Nodes: Users
  - Edges: ≥ 1 call

- **Geo-Tweets graph:**
  - Nodes: Twitter ids
  - Edges: mutually declared friendship
Social Filtering

- **Friends:**
  - Graph neighbors

- **Sharing rides with:**
  - Friends
  - Friend-of-friends
Social Filtering

- **Friends:**
  - Graph neighbors

- **Sharing rides with:**
  - Friends
  - Friend-of-friends
### Results

- **Ride-sharing parameters:**
  - Time distribution: **30 min**
  - Distance tolerance: **1 km**
  - Delay tolerance: **10 min**

<table>
<thead>
<tr>
<th>City</th>
<th>Friends only</th>
<th>Friends of friends</th>
<th>Anybody</th>
</tr>
</thead>
<tbody>
<tr>
<td>Madrid - CDR</td>
<td>1.1%</td>
<td>19% (31%)</td>
<td>53% (65%)</td>
</tr>
<tr>
<td>NY - Tweets</td>
<td>1.2%</td>
<td>8.2% (26%)</td>
<td>44% (68%)</td>
</tr>
</tbody>
</table>

Green numbers show potential of ride-sharing projected to commuters’ population.
Conclusion

- High potential based on route overlap:
  - E.g. 53% for Madrid-CDR

- Bottleneck:
  - Willingness to ride-share
  - Riding ONLY with friends is too restrictive

- Technology and building trust:
  - Riding with friends of friends: up to 31% potential.

- Other lessons:
  - Lessons from data sets
  - Spatio-temporal constraints
  - Comparisons between cities
Thank You

Blerim Cici,
Athina Markopoulou

Nikolaos Laoutarisis,
Enrique Frias-Martinez